

Muhammad Ali Norozi

Curriculum Vitae

March 2025

Address: Einar Øfstis Veg 2, 7033 Trondheim, Norway.

Phone: +47 96 23 22 79

Email: mohammad.norozi@gmail.com

WWW: www.ntnu.no/ansatte/muhammad.ali.norozi
no.linkedin.com/pub/dir/Mohammad/Norozi

Education and Qualifications

06/2023	Ped. basic Competence (1 year)	Uniped - pedagogical basic competence program of total of 200 hours, from NTNU Norway.
06/2018	Nanodegree (Deep Learning)	Nanodegree on Fundamentals of Deep Learning www.udacity.com/course/deep-learning-nanodegree-nd101
07/2019	Nanodegree (NLP)	Nanodegree on Natural Language Processing www.udacity.com/course/natural-language-processing-nanodegree-nd892
06/2013	Ph.D.	Norwegian University of Science and Technology (NTNU). Topic: Relevancy in Schema-Agnostic Environment Advisors: Adj.Assoc.Prof. Øystein Torbjørnsen, Prof. Jon Atle Gulla & Prof. Svein-Olaf Hvasshovd.
06/2008	M.Sc. (2 years)	Center of Mathematics for Application, Univeristy of Oslo (UiO).
06/2004	B.Sc. (4 years)	National University of Computers and Emerging Sciences (NUCES aka FAST), Karachi, Pakistan.

Professional Experience

06/2021-Present	Assistant Professor	NTNU, Trondheim, Norway.
01/2020-12/2023	Senior Data Scientist / Engineer	Kantega AS, Trondheim, Norway.
09/2018-12/2019	Senior Data Scientist	EVERY Norway Financial Services AS, Trondheim, Norway.
11/2013-09/2018	Data Scientist	Zedge Europe AS, Trondheim, Norway.
11/2009-04/2013	Research Fellow	Department of Computer and Information Sciences, Norwegian University of Science and Technology, Trondheim, Norway.
10/2011-06/2012	Guest Researcher	Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands.
06/2007-11/2009	Software Engineer	Fast Search and Transfer - A Microsoft Subsidiary, Oslo, Norway.
08/2005-06/2006	Software Engineer	MobileComplete aka DeviceAnywhere Inc. (acquired by Keynote) Karachi, Pakistan (Based in USA).
06/2004-08/2005	Software Engineer	Simple eSolutions Pvt. Ltd aka CapEx Baxter, Karachi, Pakistan (Based in Belgium).

Projects and responsibilities @ Kantega

- Time-series modelling of data for improving energy consumption (team-lead and data scientist).
- GPT finetuning large language model (data scientist role).
- BankID AML, data engineer and system engineer (Senior data scientist role and responsible for the data integration and data related responsibilities.).
- Loopfront AS: Object detection using ML/DL method for construction (team-lead and also advisor on the machine learning part of the Loopfront product).
- Loopfront AS: AI4Cities applied for EU funding, and achieved two phases (team-lead and the main person who drove the project ahead).

- Kantega: Azure Dataplattform, AML, Fraud detection and AI. Internaal project (data scientist/engineer role).
- SpareBank 1 Kredittkort: Machine learning modell for credit scoring for credit card applications (data scientist and data engineer role).

Projects and responsibilities @ EVRY

- Anomaly detection and prediction credit card data (planning stage, data scientist and team lead).
- ChatBot - DigitalAssistent online banking (data scientist).

Projects and responsibilities @ Zedge

- Duplicate detection in both images and ringtones (tensorflow, inception), (team-lead and data scientist).
- Improving search experience (team-lead and data scientist).
- Auto complete functionality (team-lead and data scientist).
- Emotion detection in images using DL (team-lead and data scientist).
- Demographic classification based on user profile, using storm (data scientist).
- Offensive detection using DL using caffe (data scientist).
- Related images using DL (tensorflow) (data scientist).
- Day to day analysis generating reports and results to the key stakeholders (data scientist).
- Visualisation of DAU / MAU / Disappointed users and other metrics (data scientist).
- Taking care of daily scrums in absence of VP of data (data scientist).

Teaching experience and development of teaching material

- Teaching following courses at NTNU
 - INFT2005 - Quality in and Testing of Software Systems (every spring semester since 2021 - course coordinator)
 - * Introduced active involvement of industry in the course by inviting people from industry to deliver lectures.
 - INFT2508 - Cross-platform Applications Development for Mobile Devices (every fall semester since 2021)
 - IDATT1005 - Software Engineering (every spring semester since 2021)
 - * Active participation (once a week meeting with the course coordinators) across three NTNU campuses located Ålesund, Gjøvik and Trondheim.
 - INFT2502 - Specialization Project (every fall semester since 2021 - course coordinator)
 - * Introduced industry-related projects.
 - IDATT1003 - Programming 1 (every fall semester since 2021 - course coordinator)
 - * Developed the course material from bottom-up approach to top-down (from spring 2022).
 - * Divided the course into modules (Introduction, Control Structures, Data Structures, Algorithms, Object Oriented Programming, Principles of high quality object oriented programming, and the use of AI/ML help, its consequences and NTNU regulation about it.).
 - * Identify common problems and address them on a regular basis.
 - * Changed the exam form to project.

- * Actively utilize student feedback through Mentimeter, both in one-on-one interactions and via the reference group. Adapt the lecture format based on their input, such as reducing monologues in favor of more hands-on activities and group tasks.
- * Engage in active participation through weekly meetings with course coordinators across NTNU's three campuses: Ålesund, Gjøvik, and Trondheim.
- * As part of the UniPed course on pedagogical basic competence, IDATT1003 was used as a reference course in Spring 2022 and evaluated by a group of five colleagues from various departments.
- * Utilized automated grading for programming assignments with the help of a tool called CodeGrade.
- * Running experiments by introducing elective assignments to observe the active use of AI/ML models to generate code.
- * Align the course with current and relevant advancements by incorporating day-to-day technological changes. For example, explore the impact of AI/ML-generated code on creativity and emphasize industry trends and requirements.
- IDATT2105 - Full-stack application development (every spring semester since 2021)
- IDATT2106 - Software Engineering 2 with Agile project (every spring semester since 2021)
- IDATT2900 - Bachelor thesis (every fall semester since 2021)
 - * Supervised about 14 bachelor thesis student groups (2-3 students pr. group).
 - * All of the bachelor thesis students have thesis with an external client (a project from industry).
 - * Bachelor theses focused on AI/ML related problems.
- TDMA5005 - Master thesis (every semester since 2024)
 - * So far one master thesis supervised with title: *How Can AI-Powered Dynamic Pricing Outperform Traditional Pricing Strategies in Achieving Optimal Profitability?*
- Teaching / assisting with **Operating Systems** course at NTNU - Fall 2012.
- Advisor for two groups of master students (5 students in each group) in the course **Customer Driven Project** at NTNU - Fall 2012.
- Teaching Assistant in the course **Model of Concurrency** at UiO - Fall 2006 and Fall 2007
- Teaching Assistant in the course **Discrete Mathematics** at UiO - Spring 2007.
- **Computer** Instructor at Aga Khan University, Karachi Pakistan - Fall 2003.

Selected Publications

1. M.A. Norozi, *The Contextual Features in Schema-Agnostic Environment*, Doctoral thesis, Norwegian University of Science and Technology (NTNU), Trondheim, Norway 2014.
2. M.A. Norozi and P. Arvola, *Selection Fusion in Semi-structured Retrieval*, Proceedings of the 22nd International Conference on Information and Knowledge Management (CIKM), Burlingame, CA, USA 2013.
3. M.A. Norozi and P. Arvola, *Kinship Contextualization: Utilizing the Preceding and Following Structural Elements*, Proceedings of the 36th Annual ACM SIGIR Conference, Dublin, Ireland 2013.
4. M.A. Norozi and P. Arvola, *When is the Structural Context Effective?*, Proceedings of the 13th Dutch-Belgian Information Retrieval Workshop, Delft, The Netherlands 2013.
5. M.A. Norozi, P. Arvola and A. P. de Vries, *Contextualization using Hyperlinks and Internal Hierarchical Structure of Wikipedia Documents*, Proceedings of the 21st International Conference on Information and Knowledge Management (CIKM), Maui, Hawaii, USA 2012.
6. M.A. Norozi, A. P. de Vries and P. Arvola, *Contextualization from the Bibliographic Structure*, Task Based and Aggregated Search, Proceedings of the 34th European Conference on Information Retrieval (ECIR), Barcelona, Spain, 2012.

7. M.A. Norozi, *Faster ranking using Extrapolation techniques*, International Journal of Computer Vision and Image processing, IJCVIP 2011.
8. M.A. Norozi, *Relevancy in Schema Agnostic environment*, ACM Joint Conference on Digital Libraries (JCDL) 2011, June 13-17, 2011, Ottawa Canada.
9. M.A. Norozi, *Extrapolation to Speed-up Query-dependent Link Analysis Ranking Algorithms*, Frontiers of Information Technology (FIT) 2010, December 21-23, 2010.
10. M.A. Norozi, *Information Retrieval Models and Relevancy Ranking*, M.S. thesis, Centre of Mathematics for Application, University of Oslo, 2008.

Presentations

- 10/2023 **Fine tuning – GPT and LLM models - Workshop Kantega.**
 03/2023 **Breakfast meeting: Discover technological boundaries with OpenAI and transformer-models.**
 05/2022 **Machine learning AI for dummies, with BankID AML.**
 04/2022 **User Behaviour analysis - Anomalous behaviour - JavaZone.**
 11/2019 **Learning to Rank - Trondheim Machine Learning Meetup.**

R& D Interests

- Big data & analytics.
- Machine Learning / Deep Learning.
- Relevancy in Information Retrieval / Learning to Rank.
- Large language models (LLMs).
- Algorithms and optimizations.
- Contextualization and XML / Focused Retrieval.

Awards and Honors

- Gold medallist - Best student of year 1995.
- Got on-merit Scholarship from Ministry Of Science and Technology (MOST) during my B.Sc. in a competitive exam throughout Pakistan.
- Out of 4 thousands applicants who took part in the aptitude exam to get admission at NUCES (the best institution for computer sciences) for B.Sc., I was one of top 200 students qualified.
- Participated in the division 2 cricket tournaments in Oslo, Norway from 2008 - 2009 - www.cricketforbundet.no.
- Out of 1500 submissions at CIKM 2012 and 2013, my paper got accepted for publication. Similarly at SIGIR 2013.

Software Expertise

- JAVA (thrift APIs) / dotNet / Clojure / Golang / C# / R (beginner)
- Oracle/ SQL Server/ MySQL / Hive / Hbase
- Visual Source Safe/ CVS/ SVN/ GIT
- Python (ipython), PERL, Bash scripting,
- PIG, Spark (PySpark), Kafka, Hadoop
- Jenkins (continuous deployment), Docker Marathon, Kubernetes (beginner)
- DL based - Caffe, Tensorflow, PyTorch
- Azure, AWS, Google cloud
- Good in Linux, MacOS and Windows

Languages

- Norsk (Bergenstest, fluent and use as day to day language)
- English
- Persian
- Urdu

Hobbies and Interests

- Playing cricket, watching soccer and cycling.
- Learning new things.
- Taking part in speeches and social gatherings.
- Find the best way of doing anything (optimized way).

References

References available upon request.

Muhammad Ali Norozi, PhD

Senior Data Engineer / Senior Data Scientist

Ali har lang erfaring innen datateknikk og informatikk generelt og mer spesifikt innen data science, datadrevet innsikt og maskinlæring og data engineering.

I tillegg til konsulentrollen i Kantega er Ali Universitetslektor II ved NTNU. Forskning og undervisningsarbeidet er rettet mot ulike grener av data science og generelt datateknikk. Felles for forskningsaktivitetene er at de har vært rettet mot praktiske anvendelser og reelle data.

Han har lang erfaring som datascientist, dataingeniør og systemutvikler og lang erfaring med bruk av maskinlæring (ML) og dyp læring-baserte søkeløsninger, algoritmer og modeller og prodsette dem. Han har også jobbet med alle de tre store skyteknologiene; AWS (Loopfront), Azure (Kantega og BankID) og Google Cloud (Evry). Ali har erfaring med programmeringsspråk for maskinlæring som Python og R i tillegg til objektorienterte språk som Java og C# .NET. Også har hatt brukt SQL i en god del av erfaring.

Ali har en Ph.d. i informasjonsinnhenting og datavitenskap fra NTNU 2013, en M.Sc i anvendt matematikk ved Centre of Mathematics for Application fra Universitetet i Oslo 2008 og B.Sc i informatikk og ingeniørvitenskap fra National University of Computer and Emerging Sciences også kjent som FAST i Karachi, Pakistan.

Han er lidenskapelig opptatt av dyplæring, dataanalyse, datavitenskap og engineering og programmering generelt og alltid ute etter den optimale måten å gjøre ting på.

Ali er utadventt, har godt humør og er lett å jobbe med. Han beskrives som en "komplett pakke" og har en veldig god kombinasjon av tekniske og mellommenneskelige egenskaper.

Kompetanse

Maskinlæring, Deep Learning, Datasett, Klassifisering, Søkemotor, Algoritmer, Informatikk, NLP (natural language processing), Dataanalyse, API, Learning to Rank, CNN Convolutional neural networks, Arkitektur, Design patterns, Bildebehandling, Ledelse, Mobprogrammering, ETL, Bygg og deploy-prosesser, Continuous deployment, AML, Konseptutvikling, Visualisering, Statistikk, AMS

Maskinlæring og Data Science

TensorFlow, Keras, IPython, Caffe, Apache Spark, Caffe Deep Learning, Databricks, MLflow, Azure Machine Learning

Teknologi

Solr, MVC, Hadoop, Hive, Web services, .NET, J2EE, Microservices, Google API

Programmeringsspråk

JavaScript, Java, Bash, C#, Clojure, Perl, R, Python, C# .NET

Skyteknologi/Mellomvare/Plattformer

Google Cloud, Azure, Azure DevOps, Azure Functions, Azure Keyvault, AWS, Kubernetes, Docker, Marathon Docker, Terraform, Kafka, Apache Storm, Apache HBase, Marathon, Mac OS X, Linux, Azure Active Directory, Delta Lake, Azure Databricks

Databaser

MySQL, Microsoft SQL Server, Oracle DB, SQL



Epost:

muhammad.norozi@kantega.no

Mobil: 962 32 279

Født: 1979

Roller

ML ingeniør / Systemutvikler, Senior Data Scientist og Senior Data Ingeniør

Nøkkelektspertise:

- ▣ Data ingeniør
- ▣ Systemutvikling
- ▣ Maskinlæring
- ▣ MLOps
- ▣ Databricks
- ▣ SQL
- ▣ Python
- ▣ Delta Lake

Språk:

- ▣ Norsk
- ▣ Engelsk
- ▣ Farsi
- ▣ Urdu

Verktøy

Jenkins, Git, Visual SourceSafe, CVS, Subversion, Linux/RPM, CruiseControll, Octopus

PROSJEKTERFARING

08.2021 -

NTNU: Undervisning

Ali er emneansvarlig for:

- * IDATT1001 - Programmerings 1
- * INFT2005 - Kvalitet i og testing av programvaresystemer
- * INFT2508 - Kryss-plattform applikasjonsutvikling for mobile enheter
- * INFT2502 - Fordypningsprosjekt

12.2021 - 01.2023

BankID Norge AS: BankID AML

Rolle: Analytiker, Data Engineer

Utvikling av produktet BankID AML (Anti Money Laundering). En tjeneste/plattform som enklere lar bankens kunder overholde hvitvaskingsloven. Ansvarlig for flere tilkobling til eksterne og offentlige datakilder f.eks., Brønnøysundregisterene, Danske CVR og CPR, PEP register, Finsk-folkeregister og Svenske-folkeregister gjennom Roaring APIet. Både for å opprett avtalen med de, tilkoblingen og til slutt mapping til hoved produktet. Brukte .NET for de fleste av APIene.

Har vært en sentral del av å utvikle Adverse Media Screening (AMS) løsningen med språkmodellering. Her brukte vi Python, og .NET. AMS har blitt ett eget produkt og API som skal brukes både fra BankID AML APIet og Kundesjekk AS produktet.

Dataplatteform, dataanalyse og integrering av eksterne offisielle API med BankID AML-systemet.

Analytiker

Datainnsamling, dataanalyse og gjør data søkbare.

Bruk av Databricks og Delta Lake for å forenkle og erstatte API'er som har nedetid og har vært ustabile. Systemutvikling, Continuous Integration og Continuous Deployment er en del av hverdagen. I motsetning til Data Lake ETL, har Delta lake egne tre steg løsning, med Bronze, Silver og Gold steg. I tillegg til egenskaper fra Data Lake.

Data Engineer

Den siste prosjekt som heter Adverse Media Screening. Ali har hatt hoved rolle både en som har startet prosjektet, utvikle en prototype ved hjelp av Google programmable search engine. Også senere introdusert egne språkmodell som kategoriserer en nyhetsartikkel om den hadde negative omtale om person som vi sendt inn som query.

Her brukte teamet både Python og .NET.

Kompetanseområder:

• C# • .NET • Azure DevOps • Azure Active Directory • Databricks • Azure Functions • Octopus • Bygg og deploy-prosesser
Continuous deployment • Azure Keyvault • Delta Lake • AML • Konseptutvikling • AMS • Google API

10.2020 - 03.2021

Loopfront AS: Objekt deteksjon bygg

Rolle: ML ingeniør / Systemutvikler

Loopfront er Norges største sirkulære plattform for byggematerialer og inventar. Med ombruk, resirkulering og salg, får brukte byggematerialer og inventar nytt liv og økt verdi. Deres SaaS-plattform er designet for å bidra i alle ledd i verdikjeden og forvandle avfall til ressurser. Selskapet er i rask vekst og har nylig ekspandert utenfor landegrensene med kontorer og kunder både Sverige og Tyskland.

Bildegjenkjenning ved hjelp av maskinlæringsteknikker kan erstatte en stor del av manuell data-inntasting for kundene hos Loopfront. Ved hjelp av bilder detekteres og klassifiseres ulike objekttyper som eksempelvis kontorstoler, vinduer, bord osv. Før eller uten bildegjenkjenning må brukerne tast inn mye data/ opplysninger om materialene i systemet. Vi har implementert en løsning som gjør at brukerne kan ta bilder av materialene som input til systemet og få automatisk oppslag. Scenariet er slik:

Brukeren tar et par bilder og en automatisk billedeteksjonsteknologi (basert på ML) foreslår en vare. Om bildet ikke blir gjenkjent, ber systemet brukeren ta et nytt bilde med mer fokus og klarhet. Hvis bildet ikke blir gjenkjent etter et par forsøk, ber systemet brukeren om å skrive data om varen, for eksempel kontorstol, radiator, vindu, dør, osv. Etter at enten bildet er gjenkjent eller brukeren har beskrevet om varen, vil applikasjonen foreslå varen som best samsvarer med spesifikasjonen i forrige trinn. Feltene blir autofylt og brukeren kan trykke på send-knappen.

Etter innleveringstrinnet vil varen gå gjennom kvalitetssikring fasen, som vil være enten automatisk, halvautomatisk eller helt manuelt (ikke implementert i nåværende fase).

AWS microservices ble nyttet. Det ble laget en microservice som ta imot bilder og returner kategori og label på bildene. Microservicen er i produksjon og nyttes for å gjenkjenne kategori og label på bilder.

ML ingeniør / Systemutvikler

Hovedrolle i denn oppdraget var følgende:

- Forskning på flere ideer fra state-of-the-art, eksisterende arbeid.
- Valgt ett par av de eksisterende arbeid og utvikle den videre (C# og Python).
- Produksjonssetting (Microservice og GraphQL med Azure).
- Forvaltning (endre basert på kundens behov, eller også basert på andre microservices, f.eks., om det blir endring i front-enden så endres microservice her også.)

Kompetanseområder:

• AWS • .NET • Microservices • Maskinlæring • Bildebehandling

05.2021 - 08.2021

Loopfront AS: AI4Cities

Rolle: Arkitekt, Senior systemutvikler

Loopfront er Norges største sirkulære plattform for byggematerialer og inventar. Med ombruk, resirkulering og salg, får brukte byggematerialer og inventar nytt liv og økt verdi. Deres SaaS-plattform er designet for å bidra i alle ledd i verdikjeden og forvandle avfall til ressurser. Selskapet er i rask vekst og har nylig ekspandert utenfor landegrensene med kontorer og kunder både Sverige og Tyskland.

AI4Cities er et EU-finansiert prosjekt som har som mål å hjelpe europeiske byer (deltagere) med å redusere sine klimagassutslipp ved å bruke kunstig intelligens (AI). Prosjektet er en del av EUs Horisont 2020-program og ledes av byen Helsingfors i Finland. Prosjektetets mål var å gi støtte til bedrifter og entreprenører som ønsker å utvikle AI-baserte løsninger for klimavennlige byer. Prosjektet bidra til å øke bevisstheten om fordelene ved å bruke AI for å redusere klimagassutslipp og samtidig skape forretningsmuligheter for innovative selskaper. Et spennende initiativ som bidra til å skape grønnere, sirkulær økonomi og mer bærekraftige byer i Europa ved hjelp av AI-teknologi.

Ledet søknad prosessen og dermed sendt inn søknad til AI4Cities (<https://ai4cities.eu/>) fase 2. Bruk av AI/ML i å redusere CO2 eller CO2 besparelse. Ut i fra 40 deltagere 20 har blitt med videre. Loopfront AS har blitt en del av 20 styk, og ble tildelt økonomisk støtte for å videre utvikle ideene i den tilsendte søknaden.

Arkitekt, Senior systemutvikler

Ledet stortsett hele søknaden. Skrev design dokument for prototype fase nemlig fase 2. Snakket med tredje part leverandør for material CO2 data, som cobuilder as, EPDNorge, NOBB osv.

Planlegg møter med tredje part. Snakket med folk både i Loopfront og AI4Cities.

Kompetanseområder:

• Ledelse

04.2021 - 11.2021

Kantega: Azure Dataplattform, AML, Fraud detection og AI

Rolle: Data Scientist, Data Engineer

Kantega mottok støtte fra Innovasjon Norge til å kartlegge mulighetene for ny virksomhet innen området antihvitvasking. Norske banker og andre rapporteringspliktige har betydelige krav knyttet til å avdekke hvitvasking og rapportere mistenkelig aktivitet til myndighetene.

Data Scientist

Utforsket muligheter for avansert grafanalyse for å avdekke mistenkelige transaksjoner. Satt opp infrastruktur for ML i Azure, og benyttet MLflow som rammeverk for sporbar utvikling. Infrastruktur som kode med Terraform.

Data Engineer

Innhenting av data og implementasjon av Delta lake arkitektur. Mob programmering.

Kompetanseområder:

• Azure DevOps • Azure Databricks • Azure Machine Learning • Azure Functions • Azure Keyvault • Dataanalyse • ETL • Terraform • Mobprogrammering • MLflow

01.2020 - 04.2020

SpareBank 1 Kredittkort: Maskinlæringsmodell for kredittscoring

Rolle: ML ingeniør / Systemutvikler

Kantega bisto SpareBank 1 Kredittkort med utvikling og integrasjon av en maskinlæringsbasert kredittscoringmodell - ApplicationScore. Målet med prosjektet er å utvikle en egenlaget tjeneste som kan gi bedre estimater på kredittrisiko for kredittkort søknader enn de eksterne tjenestene som brukes i dag. I tillegg til utvikling av selve modellen hadde teamet fra Kantega ansvar for eksponering av modellen mot det omliggende søknadssystemet. Maskinlæringsmodellen er nå under evaluering i produksjonssystemet.

ML ingeniør / Systemutvikler

Utviklet integrasjonsløsning for å kunne bruke maskinlæringsmodeller i søknadssystemet. Utforskning og utvikling av forskjellige prediksjons statistiske modeller. Laget API for ML tjenesten.

Kompetanseområder:

• Python • Docker • Jenkins • Shifter • Java • Maskinlæring • Arkitektur

04.2020 - 09.2020

Datadrevet Innovasjon: Forskning og kompetanseheving arbeid

Rolle: Data scientist

Kantega gjennomførte i 2020 et kartleggingsprosjekt for å forstå hvordan kunder bruker maskinlæring i dag, og hva de kommer til å satse på framover. Dette arbeidet innebar å intervju ledende eksperter som jobber i det daglige med utviklingen av en datadreven organisasjon. Arbeidet ledet fram til en rapport som ligger på kantega.ai. I arbeidet utviklet deltakerne forståelse for de forretningsmessige valgene som organisasjoner står overfor når de ønsker å bli mer datadrevne.

Data scientist

- Intervju ledende eksperter som jobber med datadreven organisasjon.
- Oppsummere og ta ut innsikt i fra intervjuene.
- Transkribere og oppsumere.

Kompetanseområder:

• Azure • Azure DevOps • MLflow • Terraform

01.2019 - 10.2019

EVERY - TietoEVERY: Conversational AI - finans assistent

Rolle: Senior Data Scientist og Senior Data Ingeniør

Brukte NLP (natural language processing) teknologier for å prøve og løse følgende:

- Intent klassifikasjon
 - Forstå brukerens intent og sammenlikne med intent fra databasen.
 - Intent klassifikasjon fra RASANLU har blitt brukt (<https://github.com/RasaHQ/rasa>)
- Stave Korreksjon for nordisk språk
 - Bruk av maskinlært systemer for eksempel (<https://github.com/google-research/bert>)
 - Bruk av vanlig løsning for eksempel: <http://norvig.com/spell-correct.html> ga bedre resultat enn ML basert.
- Word2vec for norsk
 - (<http://vectors.nlpl.eu/repository>)

Senior Data Scientist og Senior Data Ingeniør

Jobbet tett med EVRY teamet i Oslo. Ansvarlig for NLP delen av prosjektet. Laget en chatbot som kan løse enkle nettbank relatert problemer, f.eks., å fjern sperring på BankID og nettbank innloggingen.

Kompetanseområder:

• Maskinlæring • Klassifisering • NLP (natural language processing) • Java • Google Cloud • Python • Dataanalyse • Datasett • Eksperimentering

03.2017 - 06.2017

Zedge: Autofullfør funksjonalitet i søkefelt

Rolle: Data Scientist / Data Ingeniør

Ikke bare autofullfør (autocomplete) når brukere skriver søkeord, med tekst basert anbefaling men også ta hensyn til andre signaler også som f.eks, popularitet verdier (klikk, laste ned, preview osv), lokasjon basert (når man skriver 'st' i trondheim burde man se 'st. olav' i ett eller annet sted i anbefaling) og andre content basert og maskin lært basert.

Prosjektet var todelt. Første del handler om å hent riktige verdier fra logg og andre kilder. Og i andre del bruke de innsamlede og aggregate verdiene i en kjapp og responsiv API (java thrift) og server. Brukt JAVA for API-delen og Hadoo, Pig og Python for datainnhentesdelen.

Data Scientist / Data Ingeniør

Hovedrollen var å få på plass et API (med continous deploment mekanisme) og hent riktige data fra forskjellige datakilder.

Kompetanseområder:

• Java • Thrift • Python • PIG • Hadoop • Hive • Jenkins • Marathon Docker • Azkaban • Kubernetes • API

01.2016 - 05.2016

Zedge: Relaterte og dupliserte bilder

Rolle: Data Scientist / Data Ingeniør

Vi var to data scientists / data engineers i dette prosjektet. Her var problemstillingen at etter en bruker trykte på ett skjermbilde burde det kom flere andre bilder med samme mønster, samtidig brukt tidligere, farge histogram, samme objekter, osv. Her brukt vi pre-trent tensorflow inception model. Som ga oss et veldig godt resultat. En av de viktigste problemstillingene å løse i dette prosjekt var: hvordan kan man optimalisere sammenligningen. Og for å gjør det brukt vi similarity basert tre, som heter nmslib (<https://github.com/bileg/nmslib>).

I del 2 av prosjektet var problemstillingen å få fjerne dupliserte bilder. Siden det var brukergenerert content (UGC), det var ganske ofte tilfeller av duplisert bilder lastet opp av flere brukere. Igjen brukt vi tensorflow inception pre-trent med forskjellige parameter verdier og terskler.

Data Scientist / Data Ingeniør

Kompetanseområder:

• Java • Clojure • nmslib • TensorFlow • Keras • Maskinlæring • Deep Learning • Python • Jenkins • Marathon Docker • Azkaban

01.2015 - 04.2015

Zedge: Menneske hud deteksjon - eller offensiv deteksjon

Rolle: Data Scientist / Data Ingeniør

I featured side innen zedge appen, var en av de viktige reglene å ikke vis menneskehud, det betyr ikke vis menneske med mindre og uten klar. Første steg var å finne riktig trainings datasett. Dette var datasett som inneholder positivt og negativt eksempler til problemstillingen. Etter dette brukt vi Caffe deep learning bibliotek for å trene og teste. Vi oppnådde ganske gode resultater. I tidligere løsninger som ikke var basert på Deep learning var nøyaktigheten rundt 70% men etter caffe deep learning basert løsnin, fikk vi en nøyaktighet på rundt 90%.

Data Scientist / Data Ingeniør

Vi var to data scientists i dette prosjekt. Begge to jobbet med stort sett i hele systemet.

Kompetanseområder:

• Maskinlæring • Deep Learning • Caffe Deep Learning • Datasett • Java • Clojure • TensorFlow • Python • Apache HBase

06.2014 - 10.2014

Zedge: Emosjon deteksjon i bilder "skjermbilder"

Rolle: Data Scientist / Data Ingeniør

Her var problemstillingen, etter lansering av et nytt produkt "Zedge Snakk", å automatisk kategorisere bilder i flere kategorier. Til å begynne med var det å finne flere "emosjoner" i bilder: glade bilder, triste bilder osv. Her brukt jeg mest tid å komme med en riktig traineeset. Og etter den trainee, teste og validere.

Data Scientist / Data Ingeniør

Prosjektet ble i det store grad styrt av Ali, med hjelp fra backendressurser for å få på plass riktige APIer. Prodsette data innhenting (extraction), data rensing og resizing (transformation), og til slutt bruk den data for ML modellen (load). Brukt en god del av tiden for å hente data, og transformere den. Laget eget API for det, i Python plus JAVA.

Kompetanseområder:

• Klassifisering • Java • MySQL • JavaScript • Maskinlæring • Deep Learning • CNN Convolutional neural network • Gaffe

12.2013 - 03.2014

Zedge: Forbedre søkmotor i Zedge Appen

Rolle: Data Scientist / Data Ingeniør

Del 1 av prosjektet var å forbedre søkemotoren og bli kjent med hele Zedge systemet.

I Del 2 av prosjektet var planen å få inn maskinlært basert ranking funksjon. Derfor ble det brukt en "Learning to Rank" modell for å oppnå det.

Data Scientist / Data Ingeniør

Ali var ansvarlig for å drive hele prosjektet selv.

Ali har skrevet en "plugin" i solr søkemotor som henter popularitets og andre viktige signaler enn bare tekst og argumenterer den i relevance funksjon.

Kompetanseområder:

• Solr • Søkemotor • Maskinlæring • Java • PIG • Hadoop • RPM • Dataanalyse • Learning to Rank • SVM basert

11.2009 - 07.2013

NTNU: PhD Forskningsarbeid

Informasjon innhenting (information retrieval) er en av de viktigste grenene av data science og informasjonsflyt og tilgjengelighet. Forskningsarbeidet rettet mot anvendelse av data mining, prosessering og retrieval fra store mengder data med ustrukturert og strukturert tekst. Arbeidet fokuserte på å utforske teknikker som lar seg anvende i stor skala på reelle industridata, og det ble gjennomført studier med data fra både industri, og akademien. Data fra ulike kilder ble trukket ut, aggregert og sammenstilt i rike hendelseslogger. Ved å anvende data mining og statistiske metoder fikk man effektivisert den reelle verdistrømmen og identifisert avvik og flaskehalser i prosessene.

Kompetanseområder:

• Dataanalyse • Maskinlæring • SQL • Visualisering • Statistikk

01.2008 - 11.2009

Microsoft: Build

Rolle: Build Engineer

Var ansvarlig for bygging, rapportering og godkjenning av bygg prosjekter i forskjellige plattform og OS. Også laget verktøy rundt automatisk bygging og pakking av binærfiler, test og deploy.

Build Engineer

Support for bygg relatert. Godkjenne/avlyse jira saker knytte til ett bygg. Lag ny verktøy for å overføre, teste og deployet binærfiler.

Kompetanseområder:

• Java • CruiseControl • Bash • Perl • Python • .NET

05.2007 - 12.2007

Fast Search & Transfer ASA: Pluggable web-services

Rolle: Software Engineer

Laget pluggbar web-service for forskjellige tjenester, f.eks., matcher service. Og for "Did you mean" bibliotek. Legg på nye ord etter at et ord har blitt stilt mer en "k" ganger.

Software Engineer

Utviklet hele pluggable web-service i JAVA.

Kompetanseområder:

• Java • Web services • Pluggable • Design patterns • MVC

ARBEIDSERFARING

Arbeidsgiver

NTNU	07.2021 -
Kantega AS	11.2019 -
EVERY Financial Services	09.2018 - 11.2019
Zedge Europe AS	12.2013 - 09.2018
NTNU	11.2009 - 06.2013
Centrum Wiskunde & Informatica (CWI), Amsterdam	10.2011 - 06.2012
Fast Search & Transfer	06.2007 - 09.2009
Universitetet i Oslo	08.2006 - 12.2007
MobileComplete aka DeviceAnywhere Inc	08.2005 - 06.2006
Simple eSolutions Pvt. Ltd	06.2004 - 08.2005

Periode

Stilling

Universitetslektor II
Senior konsulent
Senior Data Scientist
Data Scientist
Stipendiat
Gjesteforsker

Softwareutvikler
Læringsassistent
Software utvikler
Software utvikler

Referanser

Referanseperson

Knut Ola Hellan

Øystein Torbjørnsen

Jon Atle Gulla
Geir Dahl

Referanse

CTO CompanyBook og Freelance AI og data management rådgiver
knut@companybook.no
mobil: 91639568.

Var team lead i data science teamet hos zedge fram til 2015, jobbet ca 3 år med han.
Senior Partner Program Manager, Microsoft Development Centre Norway.
oytor@oytor.com
mobil: 90892977.
Hovedveileder i doktorgraden fra 2009 til 2013.

Professor og bi-veileder på NTNU
Professor og veileder på Matematisk institut UiO.

UTDANNELSE

Skole

NTNU	2009 - 2013
Universitet i Oslo, Center of Mathematics for Application (CMA)	2006 - 2008
National University of Computers and Emerging Sciences (NUCES aka FAST), Karachi	2000 - 2004

Periode

Retning

Ph.D. informasjonsinnhenting og datavitenskap
Master

Bachelor

Kurs og kompetanseheving

Dato	Kurs	Arrangør
02.2020	Mobbprogrammering	Kantega / Woody Zuill
07.2019	Nanodegree: Natural Language Processing	Udacity.com
04.2018	Nanodegree: Fundamentals of Deep Learning	Udacity.com

PUBLIKASJONER

06.2014	The Contextual Features in Schema-Agnostic Environment. M.A. Norozi, Doctoral thesis, Norwegian University of Science and Technology (NTNU), Trondheim, Norway 2014. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/253650
06.2013	Selection Fusion in Semi-structured Retrieval. M.A. Norozi and P. Arvola. Proceedings of the 22nd International Conference on Information and Knowledge Management (CIKM), San Fransisco, USA 2013. https://dl.acm.org/doi/10.1145/2505515.2505686
08.2013	Kinship Contextualization: Utilizing the Preceding and Following Structural Elements. M.A. Norozi and P. Arvola. Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval (SIGIR), Dublin, Ireland 2013. https://dl.acm.org/doi/10.1145/2484028.2484111
06.2013	When is the Structural Context Effective?. M.A. Norozi and P. Arvola. Proceedings of the Dutch- Belgian Information Retrieval Workshop (DIR), Delft, The Netherlands 2013. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2395979?locale-attribute=en
06.2012	Contextualization using Hyperlinks and Internal Hierarchical Structure of Wikipedia Documents. M.A. Norozi, P. Arvola and A. P. de Vries. Proceedings of the 21st International Conference on Information and Knowledge Management (CIKM), Maui, Hawaii, USA 2012. https://dl.acm.org/doi/10.1145/2396761.2396855
2012	Contextualization from the Bibliographic Structure. M.A. Norozi, A. P. de Vries and P. Arvola. Proceedings of the 34th European Conference on Information Retrieval (ECIR), Barcelona, Spain, 2012. https://www.semanticscholar.org/paper/Contextualization-from-the-Bibliographic-Structure-Norozi-Vries/19231dcaee301e08aec5d2d2be810e3ec591370d
2011	Faster Ranking Using Extrapolation Techniques. M.A. Norozi. International Journal of Computer Vision and Image processing, IJCVIP 2011. https://www.igi-global.com/article/faster-ranking-using-extrapolation-techniques/59877
06.2011	Relevancy in Schema Agnostic Environment. M.A. Norozi. ACM Joint Conference on Digital Libraries (JCDL) 2011, June 13- 17, 2011, Ottawa Canada. https://www.researchgate.net/publication/243464712_Relevancy_in_Schema_Agnostic_Environment
12.2010	Extrapolation to Speed-up Query-dependent Link Analysis Ranking Algorithms. M.A. Norozi. Frontiers of Information Technology (FIT) 2010, December 21st - 23th, 2010. https://dl.acm.org/doi/abs/10.1145/1943628.1943630
2008	Information Retrieval Models and Relevancy Ranking. M.A. Norozi. M.Sc. Thesis, Centre of Mathematics for Application, University of Oslo, 2008. https://www.duo.uio.no/handle/10852/10761

PRESENTASJONER

Dato	Presentasjon
11.2019	Learning to Rank. Ga en oversikt over området L_Earning TO Rank (LETOR). Relevans-scoring er kjernen i informasjonsinnhenting. Det handler om hvor relevant informasjon i databaser til brukernes sporing, informasjon er tilgjengelige dokumenter, som kan være bilder, artikler, videoer, lydfiler, alt. Relevanse-scoring er å sortere dem alt etter deres grad av relevans til brukernes stilte sporing eller søk, i det liste den første er den mest relevante. Hva har LETOR å gjøre med maskinlæring? Hvorfor lære og relevanse? Hva har relevanse å gjøre med læring og hva har læring "gitt" for rangeringsproblemet? Relevante og eksisterende forskningsarbeid i område og kategoriseringen i akademia om å lære å ranke og noen eksempler og motivasjoner fra den virkelige verden. Beregnet publikum var: dataforskere og programvareutviklere. Dette møtet var en del av en serie læringsøkter for AI-landsbyer som blir arrangert hver siste torsdag i måneden (unntatt helligdager).